Running Head: MEASUREMENT BIAS AND ERROR CORRECTION

Measurement bias and error correction in a two-stage estimation for multilevel IRT models

Xue Zhang[1]

China Institue of Rural Education Development, Northeast Normal University, China

Chun Wang[2]

College of Education, University of Washington, USA

Correspondence concerning this manuscript should be addressed to Xue Zhang at:

China Institute of Rural Education Development,
Northeast Normal University,
5268 Renmin Street,
Changchun, Jilin Province, 130024
e-mail: zhangx815@nenu.edu.cn
phone: 86-13604412221

The data that support the findings of this study are available from NCES. Restrictions apply to the availability of these data, which were used under license for this study. Data are avilable from https://nces.ed.gov/surveys/nels88/ with the permission of NCES. The Research was partly supported by IES R305D170042 and National Natural Science Foundation of China (Grand 12001092).

Citation:
Zhang, X., & Wang, C. (2021). Measurement bias and error correction in a two-stage estimation for multilevel IRT models. *British Journal of Mathematical and Statistical Psychology*. DOI: 10.1111/bmsp.12233

[1] Xue Zhang, 5268 Renmin Street, Changchun, Jilin Province, China, 130024.
[2] Chun Wang, 312E Miller Hall, Box 353600, Seattle, WA, USA, 98195-3600.

Abstract

Among current state-of-art estimation methods for multilevel IRT models, the two-stage divide-and-conquer strategy has practical advantages, such as clearer definition of factors, convenience for secondary data analysis, convenience for model calibration and fit evaluation, and avoidance of improper solutions. However, various studies have shown that, under the two-stage framework, ignoring measurement error in the dependent variable in stage II leads to incorrect statistical inferences. To this end, we proposed a novel method to correct both measurement bias and measurement error of latent trait estimates from stage I in the stage II estimation. In this article, the HO-IRT model was considered as the measurement model, and a linear mixed effects model on overall (i.e., higher-order) abilities was considered as the structural model. The performance of the proposed correction method was illustrated and compared via a simulation study and a real data example using the National Educational Longitudinal Survey data (NELS 88). Results indicated that structural parameters could be recovered better after correcting measurement biases and errors.

Keywords: HO-IRT models; two-stage estimation; measurement bias; measurement error

Measurement Bias and Error Correction in a Two-stage Estimation for Multilevel IRT Models

1. Introduction

Item response theory (IRT) has been extensively used to analyze survey, instrument, questionnaire, or test data, and the resulting IRT-based scaled scores (i.e., θ) have been used as proxies of latent traits, such as academic achievement, skill proficiency, personality, or quality of life, among others. Once θ is obtained from IRT analyses, they are often used as covariates or dependent variables in follow-up analyses. In educational measurement context for instance, θ has been used as dependent variables in two sample t-test, multiple regression (Goldhaber & Brewer, 1997; Nussbaum, Hamilton, & Snow, 1997), analysis of variance (ANOVA, Cohen, Bottge, & Wells, 2001), hierarchical linear modeling (Bacharach, Baumeister, & Furr, 2003), just to name a few. Oftentimes, practitioners use θ in statistical analyses as if they were "true" values without measurement biases or measurement errors. Ignoring measurement error will inevitably diminish the statistical power of impact studies, yield inconsistent or biased estimates of model parameters (Lu, Thomas, & Zumbo, 2005), and weaken presumed relationships among different variables affecting outcomes (Skrondal & Rabe-Hesketh, 2004).

The issues that emerge when one tries to use latent variables as outcomes in regression analysis can be addressed by taking a multilevel perspective on item response modeling (Adams, Wilson, & Wu, 1997). Hierarchical models have been proven useful for solving the technical problems that arise when traditional approaches and models are applied to nested data, such as students nested within classrooms or repeated measures nested within persons (Raudenbush & Bryk, 2002). Within the framework of multilevel or hierarchical models, the IRT model is placed at the lowest level as a within-subject measurement model and the

student population distribution is typically treated as a between-subject structural model. The multilevel IRT model makes it possible to simultaneously estimate the item and ability parameters and the structural multilevel model parameters (e.g., Adams, et al., 1997; Kamata, 2001; Pastor, 2003). Therefore, measurement error in the estimated abilities is well handled in estimating the multilevel parameters (Adam, et al., 1997). This approach is called the unified one-stage approach. The potential advantages of multilevel models include (1) direct estimation of population parameters from the item responses obviates the problem of the bias introduced by two-step estimation; (2) the use of student-level variables can lead to increased precision in estimation of item parameters (Mislevy, 1987; Mislevy & Sheehan, 1989) and increased precision in the estimation of person parameters.

The current state-of-art one-stage estimation methods include, for instance, the generalized linear and nonlinear methodologies described in De Boeck and Wilson (2004), the generalized linear latent and mixed modeling (GLLAMM) framework of Skrondal and Rabe-Hesketh (2004), adaptive Gaussian quadrature method (Pinheiro & Bates, 1995), stochastic expectation-maximization algorithm (von Davier & Sinharay, 2007), limited-information weighted least squares, graphic models approach (Rijmen, Vansteelandt, & De Boeck, 2008), as well as Bayesian methodology of Lee and Song (2003) such as the Gibbs sampler and Markov chain Monte Carlo (MCMC). Although all these methods have proven to work well in respective studies, the estimation could be slow (Rabe-Hesketch & Skrondal, 2008) or suffer from Heywood cases or convergence issues (Anderson & Gerbing, 1984), especially when the number of random effects is prohibitively large. Moreover, the one-stage approach is also known to have other disadvantages. First, a conceptual challenge may arise, which Burt (1976) referred to as “interpretational confounding”, meaning that the definition of the latent construct in the measurement model may change after altering the structural model such as adding or removing covariates. Second, one misspecification in the model may

propagate through many levels and influence the performance of the whole model (Bollen, 1996, Croon, 2002). As a result, the whole model is always refitted even if only one part of the model is changed, and practically, this can make the estimation computationally demanding. Instead, stepwise methods avoid the problems of the one-step approach by separating the estimation of the different parts of the model into distinct steps of analysis. Hence, the stepwise methods enjoy some nice features such as easier model calibration and goodness-of-fit analysis, and avoidance of improper parameter estimation (Wang, Xu, & Zhang, 2019). Moreover, stepwise approach is compatible with secondary data analysis. For instance, a large-scale survey such as NAEP, usually hundreds of test items and educational, demographic, and attitudinal variables are included, such that droves of descriptive statistics, multiple regression analyses, and SEM models might be entertained. In this case, neither carrying out all these analyses nor providing sufficient statistics for them is feasible. Oftentimes, these survey data provide either item parameters, or estimated θs along with their standard errors. It is therefore desirable for substantive researchers to perform statistical analysis with this available information.

Disregarding measurement errors in latent variables in a naïve stepwise approach, although quite prevalent in practice, is statistically suboptimal. Instead, a number of researches in different disciplines and contexts have been devoted to developing viable stepwise approaches that properly address measurement issues. For instance, in the context of classical test theory, Lord and Novick (1968) proposed a correction for attenuation to account for unreliability of measurement scales. In the context of factor analysis, issues with using factor scores in linear regression were studied by Tucker (1971), Skrondal and Laake (2001), Croon (2002), among others. Specifically, Skrondal and Laake (2001) took what Lu and Thomas (2008) claimed as a “bias avoidance” approach. That is, they showed a combination of Bartlett factor scores (Bartlett, 1937) for dependent variables and regression factor scores

for independent variables yield consistent estimates of the parameters of a set of linear latent equations. Their approach, originally proposed for continues observed variables, was extended by Lu and Thomas (2008) to binary or ordinal observed variables. Croon (2002), on the other hand, proposed a bias correct approach to adjust the variance-covariance of factors scores so as to remove bias in regression parameter estimates. In the context of structural equation modeling, measurement error challenges have been extensively discussed by Bollen (1989). And in statistics, there is a large body of literature documenting difficulties and potential solutions for "errors in variables" in linear models by Fuller (1987) and for nonlinear models by Carroll, Ruppert, and Stefanski (1995).

While the previous literature focuses either on continuous observed variables predominantly, or a single-level IRT model as a measurement model, or regression model as a structural model, in this paper, we will explore the two-stage approach for a specific type of multilevel IRT models, namely, the longitudinal higher-order IRT model (L-HO-IRT; de la Torre & Song, 2009; Huang, 2015; Wang & Nydick, 2019). There are two unique features of the model that warrants new methods to handle measurement errors. First, the measurement model is a higher-order IRT (HO-IRT) model, hence measurement errors exist for both lower-order and higher-order factors and they are heterogenous in the population. Second, the structural model is a latent growth curve model, known as a special case of general mixed effects models. The random effects in this model add further complication to estimation in stage II and therefore Croon (2002)'s corrective approach cannot easily apply.

The L-HO-IRT model is of particular interest because the HO-IRT model captures the hierarchical nature of multidimensional latent traits, which are widely seen in psychology over a variety of domains, such as cognitive ability (Murray & Johnson, 2013), quality of life (Gotay, Blaine, Haynes, Holup, & Pagano, 2002; Chen, West, & Sousa, 2006), intelligence (Golay & Lecerf, 2011), personality (DeYoung, 2006), or psychological well-being (Hills &

Argyle, 2002), etc. In this model, it is typically hypothesized that the general construct is comprised of several highly related group specific factors, each of which is measured by multiple indicators, referred to as items. For example, in many educational assessments, one is often required to report both overall proficiency for accountability purposes as well as domain-specific proficiency for diagnostic purposes. The longitudinal HO-IRT model (L-HO-IRT) extends HO-IRT by imposing a structural latent growth curve (LGC) model on the general factor over time, and the growth on the lower-order factors is therefore determined through their linkage to the general factors. Although the LGC model is considered in the present study, given that LGC is a special case of linear mixed effects (LME) models, the methods discussed therein could be applied to other LME models as well with little modification needed.

This article adopts and extends the two-stage approach in Wang et al. (2019). In stage I, the maximum a posteriori (MAP) is used to obtain the point estimates of latent traits (Wang, 2015). However, MAP is biased. Therefore, in stage II, both measurement *bias* and measurement *error* in the dependent variable need to be accounted for simultaneously when fitting an LME model. To this end, we propose a new corrective method to deal with heterogeneous measurement bias and error. Other comparison methods include naïve two-stage estimation, moment estimation method (MEM), and one-stage MCMC algorithm. Note that the MEM bears some resemblance to the Croon (2002)'s corrective approach in that instead of computing the sample variance-covariance matrix of the *estimated* latent factors in stage I, we can directly estimate the variance-covariance matrix of the *true* latent factors. The performance of these methods will be evaluated thoroughly in both the simulation study and a real data example.

The reminder of this article is organized as follows. In section 2, we introduce a correction model that takes into consideration both measurement biases and errors. This

correction model is then used for stage II estimation and in section 3, where we propose two new two-stage estimation methods. Simulation study and a real data analysis are provided in sections 4 & 5, respectively. We end with the conclusions in section 6. The detailed analytical derivations are provided in the Appendix.

2. Model Description

The model is comprised of two levels: the measurement model to explain the multi-unidimensional and hierarchical structure of latent traits and the structural model. Moreover, in the two-stage estimation, a correction model is introduced in stage II to correct the measurement bias and error simultaneously.

Measurement model

Let y_{ijt} denote the dichotomous response of examinee *i* (*i* = 1, …, *N*) on item *j* (*j* = 1, …, *J*) that belongs to domain *k* (*k* = 1, …, *K*) at group *t* (*t* = 1, …, *T*; e.g., gender, time, school, or country), then the probability of a correct response is given by

$$p\left(y_{ijkt}=1\middle|a_{jkt},b_{jt},c_{jt},\theta_{ikt}\right)=c_{jt}+\frac{1-c_{jt}}{1+\exp[-Da_{jkt}(\theta_{ikt}-b_{jt})]}, \tag{1}$$

$$\theta_{ikt}=\lambda_k\xi_{it}+\varepsilon_{ikt}, \tag{2}$$

where a_{jkt} is the discrimination parameter of item *j* belong to domain *k* at group *t*, b_{jt} and c_{jt} are the difficulty and guessing parameters of item *j* at group *t*, θ_{ikt} is the *k*th domain-specific ability of examinee *i* at group *t*, and *D* is the scaling constant, which is usually set to 1.7. λ_k denotes the regression coefficient, ξ_{it} is the overall ability of examinee *i* at group *t*, and ε_{ikt} denotes the *k*th domain-specific residual of examinee *i* at group *t*, which follows a normal distribution, $N(0,1-\lambda_k^2)$. Interested readers can refer to de la Torre and Song (2009) for a full description of the family of HO-IRT models.

The HO-IRT model is an extension of the MIRT model by introducing a higher-order ability/factor to handle ability/factor hierarchy. For an item conforming to MIRT, it can have

one (i.e., between-item multidimensional structure) or multiple (i.e., within-item multidimensional structure) discrimination parameters, but it will only have one difficulty parameter and one guessing parameter as these two type of parameters are "item" level, not item-by-domain level[3].

By plugging Equation 2 into Equation 1, the hierarchical model can be written as a one-level model as follows (Huang, Chen, & Wang, 2012),

$$\begin{aligned} p_{ijkt} &= p\left(y_{ijkt}=1 \middle| a_{jkt}, b_{jt}, c_{jt}, \lambda_k, \xi_{it}, \varepsilon_{ikt}\right) \\ &= c_{jt} + \frac{1-c_{jt}}{1+\exp[-Da_{jkt}(\lambda_k \xi_{it} + \varepsilon_{ikt} - b_{jt})]}. \end{aligned} \quad (3)$$

Given the known item parameters and regression coefficients, both ξ_{it} and ε_{ikt} could be obtained via MAP in stage I.

Please note that we have subscript *t* in above Equations 1-3 to show that item parameters and residuals can differ across groups or time. But to establish a common scale, measurement invariance needs to be satisfied. For HO-IRT model across groups or time, there are seven levels of invariance (from least restrictive to most restrictive) that need to be checked: (1) configural invariance (i.e., an unrestricted baseline model), (2) invariance of first-order factor loadings[4], (3) invariance of second-order factor loadings[5], (4) invariance of intercepts of measured variables[6], (5) invariance of intercepts of first-order latent factors, (6) invariance of disturbances of first-order factors, and (7) invariance of residual variance of observed variables. As checking invariance is beyond the scope of the current paper, for more details, please refer to Chen, Sousa, and West (2005). In this article, item parameters are fixed to be the same over time (satisfying invariance 1, 2, 4), all the higher-order factor loadings are

[3] Indeed, due to the compensatory nature of a typical MIRT model, having multiple difficulty parameters will render the model non-identifiable. And for the guessing parameter, it is defined as the probability of guessing the item right when someone has infinitely low abilities on relevant domains. Hence, it is an item level parameter as well.

[4] This level of invariance is the same as invariance of item discrimination parameters.

[5] This level of invariance is the same as invariance of λ_k's.

[6] In IRT terminology, this level of invariance is the same as invariance of item difficulty parameters.

constrained to be the same over time (satisfying invariance 3). Invariance 5 and 7 are not applicable as the intercept are fixed at 0 and residuals are absorbed in the IRT model itself[7]. The correlation between higher-order abilities is freely estimated using the LME model.

Structural model

In the structural model level, we will focus specifically on the linear mixed effects on the overall ability, $\boldsymbol{\xi}_i$, which can be expressed as

$$\boldsymbol{\xi}_i = \boldsymbol{X\beta} + \boldsymbol{Zu_i} + \boldsymbol{e}_i \ , \tag{4}$$

where $\boldsymbol{\beta}$ is a *w*-dimensional vector of fixed effects, $\boldsymbol{u}_i$ is an examinee-specific *v*-dimensional vector of random effects. The matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ are design matrices of size $T \times w$ and $T \times v$ for the fixed and random effects, respectively, where *T* denotes the total number of groups. $\boldsymbol{e}_i$ is a noise term that follows an independent and identical normal distribution (i.e., $\boldsymbol{e}_i \sim N(0, \sigma^2)$). The random effects, $\boldsymbol{u}_i$, are typically assumed to follow a multivariate normal distribution[8] with mean zero and covariance $\boldsymbol{\Sigma}_u = \{\tau_{hl}\}$, where $h, l = 1, \ldots, v$.

Correction model

As the latent variables (i.e., overall abilities $\boldsymbol{\xi}$ and domain-specific residuals $\boldsymbol{\varepsilon}$) are measured with bias and error in stage I, one can only observe $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\varepsilon}}$. To correct the measurement bias and error simultaneously in stage II estimation, a measurement error model needs to be introduced to account for heterogeneous biases and errors as follows,

$$\hat{\boldsymbol{\xi}}_i = \boldsymbol{\xi}_i + \boldsymbol{\delta}_{\xi i} \ \text{ and } \ \hat{\boldsymbol{\varepsilon}}_i = \boldsymbol{\varepsilon}_i + \boldsymbol{\delta}_{\varepsilon i} \ , \tag{5}$$

where $\hat{\boldsymbol{\xi}}_i$ and $\hat{\boldsymbol{\varepsilon}}_i$ are the point estimates of $\boldsymbol{\xi}_i$ and $\boldsymbol{\varepsilon}_i$ from stage I, respectively, and

[7] Based on the equivalency between the IRT model and factor analysis model (Kamata & Bauer,2008; Takane & de Leeuw, 1987), if writing the IRT model as a factor analysis model and using a *conditional* parameterization, one can test the equivalency of residuals over time.

[8] Generally, if the random effects follow other distribution (e.g., *t* distribution), the MCEM algorithm (proposed in next section) should be used in stage II.

$\delta_{\xi it}$ and $\delta_{\varepsilon ikt}$ are the noise terms of ξ_{it} and ε_{ikt}, respectively. We assume the noise term follows a multivariate normal distribution with blocked diagonal covariance matrix as follows,

$$\begin{bmatrix} \delta_{\xi i1} \\ \delta_{\epsilon i11} \\ \vdots \\ \delta_{\epsilon iK1} \\ \vdots \\ \delta_{\xi iT} \\ \delta_{\epsilon i1T} \\ \vdots \\ \delta_{\epsilon iKT} \end{bmatrix} \sim MVN \left\{ \hat{B}_i = \begin{bmatrix} \hat{B}_{\xi i1} \\ \hat{B}_{\epsilon i11} \\ \vdots \\ \hat{B}_{\epsilon iK1} \\ \vdots \\ \hat{B}_{\xi iT} \\ \hat{B}_{\epsilon i1T} \\ \vdots \\ \hat{B}_{\epsilon iKT} \end{bmatrix}, \hat{\boldsymbol{\Phi}}_i = \begin{bmatrix} \hat{\sigma}^2_{\xi i1} & \hat{\sigma}_{\xi\varepsilon i11} & \cdots & \hat{\sigma}_{\xi\varepsilon iK1} & & & & & & \\ \hat{\sigma}_{\xi\varepsilon i11} & \hat{\sigma}^2_{\varepsilon i11} & \cdots & \hat{\sigma}_{\varepsilon i1,K1} & & & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & & & \\ \hat{\sigma}_{\xi\varepsilon iK1} & \hat{\sigma}_{\varepsilon i1,K1} & \cdots & \hat{\sigma}^2_{\varepsilon iK1} & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & \hat{\sigma}^2_{\xi iT} & \hat{\sigma}_{\xi\varepsilon i1T} & \cdots & \hat{\sigma}_{\xi\varepsilon iKT} \\ & & & & & \hat{\sigma}_{\xi\varepsilon i1T} & \hat{\sigma}^2_{\varepsilon i1T} & \cdots & \hat{\sigma}_{\varepsilon i1,KT} \\ & & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & & \hat{\sigma}_{\xi\varepsilon iKT} & \hat{\sigma}_{\varepsilon i1,KT} & \cdots & \hat{\sigma}^2_{\varepsilon iKT} \end{bmatrix} \right\} \quad . (6)$$

In Equation 6, $\hat{\boldsymbol{B}}_i$ is a $T(K+1)$-dimensional vector of measurement biases of overall ability and residuals of examinee *i*, and $\hat{\boldsymbol{\Phi}}_i$ is the corresponding $T(K+1) \times T(K+1)$ measurement error covariance matrix. The measurement error correction model considered in Wang et al. (2019) can be viewed as a specific case of the general model introduced above, because not only did they not consider the higher-order measurement model, but they also assumed $\hat{\boldsymbol{B}}_i = \mathbf{0}$, ignoring the measurement biases of latent trait estimates from stage I.

3. Estimation Methods

3.1 Consequences of ignoring measurement biases and errors

Before introducing the two-stage methods for dealing with measurement bias and error challenges in stage II estimation, we first discuss the consequences of ignoring measurement biases and errors. When combining the measurement error model in Equation 5 with the structural LME in Equation 4, we have

$$\hat{\boldsymbol{\xi}}_i = \boldsymbol{X\beta} + \boldsymbol{Zu}_i + (\boldsymbol{e}_i + \boldsymbol{\delta}_{\xi i}) \quad , \tag{7}$$

whereas the naïve method uses

$$\hat{\boldsymbol{\xi}}_i = \boldsymbol{X\beta} + \boldsymbol{Zu}_i + \boldsymbol{e}_i. \quad (8)$$

The known results for $\hat{\boldsymbol{\beta}}$ from Equation 8 is $\hat{\boldsymbol{\beta}} = (\dot{\boldsymbol{X}}^T\dot{\boldsymbol{V}}^{-1}\dot{\boldsymbol{X}})^{-1}\dot{\boldsymbol{X}}^T\dot{\boldsymbol{V}}^{-1}\hat{\boldsymbol{\xi}}$, where $\hat{\boldsymbol{\xi}}$ is a *NT*-by-1 vector stacking all $\hat{\boldsymbol{\xi}}_i$'s vertically, $\dot{\boldsymbol{X}} = 1_{N\times 1} \otimes \boldsymbol{X}_{T\times 2}$, $\dot{\boldsymbol{V}} = \boldsymbol{I}_N \otimes \boldsymbol{V} \equiv \boldsymbol{I}_N \otimes (\boldsymbol{Z}\hat{\boldsymbol{\Sigma}}_u\boldsymbol{Z}^T + \hat{\sigma}^2\boldsymbol{I}_T)$. It can be easily shown that when $\hat{\boldsymbol{\xi}}_i$ is unbiased, then $\hat{\boldsymbol{\beta}}$ is also unbiased, i.e., E($\hat{\boldsymbol{\beta}}$)=$\boldsymbol{\beta}$. Otherwise, $\hat{\boldsymbol{\beta}}$ is biased. This implies that ignoring the measurement error does not bias the fixed effect coefficient, but ignoring the measurement bias does. However, the standard error of $\hat{\boldsymbol{\beta}}$ depends on the square root of $(\dot{\boldsymbol{X}}\dot{\boldsymbol{V}}^{-1}\dot{\boldsymbol{X}}^T)^{-1}$, or equivalently, $\frac{1}{N}(\dot{\boldsymbol{X}}^T\dot{\boldsymbol{V}}^{-1}\dot{\boldsymbol{X}})^{-1}$. Without measurement error, we have

$$\boldsymbol{V} = \boldsymbol{Z}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}\boldsymbol{Z}^T + \hat{\boldsymbol{\sigma}}^2\boldsymbol{I}_T, \quad (9)$$

whereas with measurement error, we have

$$\boldsymbol{V} = \boldsymbol{Z}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}\boldsymbol{Z}^T + \hat{\boldsymbol{\sigma}}^2\boldsymbol{I}_T + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}. \quad (10)$$

Hence, the naïve method intrinsically absorbs measurement error as part of the residual error, yielding inflated standard error for $\hat{\boldsymbol{\beta}}$. Furthermore, due to the measurement error-covariance matrix in Equation 10, the residual variance is likely over-estimated and random effects covariance matrix tends to be under-estimated, and ignoring the measurement bias does not influence them. Simple correction is available if $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ is a constant, however, in the scenario discussed in this paper, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ differs per person per time point.

3.2 Two-stage methods

Assuming item parameters and regressive coefficient (i.e., $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{\lambda}$) are known or pre-calibrated from stage I estimation, let $\boldsymbol{\Psi} = \left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma}_u\right)$ denote the set of structural parameters of interest. The existing methods (i.e., naïve method, moment estimation method and Wang et al.'s methods) will be revisited briefly following the proposed methods.

After correcting measurement bias and error, the marginal likelihood of $\boldsymbol{\Psi}$ can be

expressed as

$$\begin{aligned} L(\boldsymbol{\Psi}|\boldsymbol{X},\boldsymbol{Z}) &= \int L(\boldsymbol{\Psi}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\xi},\boldsymbol{\varepsilon},\boldsymbol{u})d\boldsymbol{\xi}d\boldsymbol{\varepsilon}d\boldsymbol{u} \\ &= \prod_i \int p\left(\boldsymbol{\xi}_i \middle| \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\boldsymbol{u}_i,\sigma^2\boldsymbol{I}_T\right) p\left(\boldsymbol{u}_i|0,\boldsymbol{\Sigma}_u\right) p\left(\hat{\boldsymbol{\xi}}_i,\hat{\boldsymbol{\varepsilon}}_i \middle| \boldsymbol{\xi}_i,\boldsymbol{\varepsilon}_i,\hat{\boldsymbol{B}}_i,\hat{\boldsymbol{\Phi}}_i\right) d\boldsymbol{\xi}d\boldsymbol{\varepsilon}d\boldsymbol{u} \\ &\propto \prod_i \int p\left(\boldsymbol{\xi}_i \middle| \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\boldsymbol{u}_i,\sigma^2\boldsymbol{I}_T\right) p\left(\boldsymbol{u}_i|0,\boldsymbol{\Sigma}_u\right) p\left(\hat{\boldsymbol{\xi}}_i \middle| \boldsymbol{\xi}_i,\hat{\boldsymbol{B}}_{\xi i},\hat{\boldsymbol{\Sigma}}_{\xi i}\right) d\boldsymbol{\xi}d\boldsymbol{u} \end{aligned}, \quad (11)$$

where $\hat{\boldsymbol{B}}_{\xi i}$ is the measurement bias of $\boldsymbol{\xi}_i$, and $\hat{\boldsymbol{\Sigma}}_{\xi i}$ is the measurement error covariance matrix of $\boldsymbol{\xi}_i$, which are presented in Equation 6, i.e.,

$$\hat{\boldsymbol{\Sigma}}_{\xi i} = \begin{bmatrix} \hat{\sigma}^2_{\xi i1} & \cdots & \hat{\sigma}_{\xi i1T} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{\xi i1T} & \cdots & \hat{\sigma}^2_{\xi iT} \end{bmatrix}.$$

Then, estimating the structural parameters, $\boldsymbol{\Psi}$, boils down to maximizing Equation 11. If the measurement bias and error and the random effect term satisfy the (multivariate) normal assumption, the marginalized maximum likelihood estimator (MMLE) of $\boldsymbol{\Psi}$ has a closed form. Otherwise, the Monte Carlo Expectation-Maximization (MCEM) algorithm should be used with the importance sampling. The detailed analytical derivations are given below.

Marginalized maximum likelihood estimation (MMLE)

Assume $p\left(\hat{\boldsymbol{\xi}}_i \middle| \boldsymbol{\xi}_i,\hat{\boldsymbol{B}}_{\xi i},\hat{\boldsymbol{\Sigma}}_{\xi i}\right)$ is a multivariate normal probability density function (PDF), then we can obtain $p\left(\boldsymbol{\xi}_i \middle| \boldsymbol{\beta},\boldsymbol{\Sigma}\right)$, which is a multivariate normal PDF with mean $\boldsymbol{X}\boldsymbol{\beta}$ and variance-covariance matrix $\boldsymbol{\sigma}^2\boldsymbol{I}_T + \boldsymbol{\Sigma}_u$, and $p\left(\hat{\boldsymbol{\xi}}_i \middle| \boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma}_u,\hat{\boldsymbol{B}}_{\xi i},\hat{\boldsymbol{\Sigma}}_{\xi i}\right)$, which is a multivariate normal PDF with mean $\boldsymbol{X}\boldsymbol{\beta} + \hat{\boldsymbol{B}}_{\xi i}$ and variance-covariance matrix $\boldsymbol{\sigma}^2\boldsymbol{I}_T + \boldsymbol{\Sigma}_u + \hat{\boldsymbol{\Sigma}}_{\xi i}$.

Overall, we can write the log marginal likelihood in a close-form as follows,

$$\begin{aligned} l(\boldsymbol{\Psi}) &= \log L(\boldsymbol{\Psi}) \propto \sum_i \log p\left(\hat{\boldsymbol{\xi}}_i \middle| \boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma}_u,\hat{\boldsymbol{B}}_{\xi i},\hat{\boldsymbol{\Sigma}}_{\xi i}\right) \\ &\propto -\log\left|\sigma^2\boldsymbol{I}_T + Z\boldsymbol{\Sigma}_u Z^T + \hat{\boldsymbol{\Sigma}}_{\xi i}\right| - (\hat{\boldsymbol{\xi}}_i - \boldsymbol{X}\boldsymbol{\beta} - \hat{\mathbf{B}}_{\xi i})^T(\sigma^2\boldsymbol{I}_T + Z\boldsymbol{\Sigma}_u Z^T + \hat{\boldsymbol{\Sigma}}_{\xi i})^{-1}(\hat{\boldsymbol{\xi}}_i - \boldsymbol{X}\boldsymbol{\beta} - \hat{\mathbf{B}}_{\xi i}) \end{aligned} \quad .(12)$$

To maximize Equation 12, the "optim" function in "stats" library of R is used. To obtain

the optimization, the L-BGFS-B method is used with box constraints: -1000< β_0, β_1<1000, 0.001<σ_{u_0}, σ_{u_1}<5, -.99<ρ<0.99, and 0.001<σ^2<1000[9]. The initial values are all set to be 0.1. In this study, the parameter point estimates are the output from the function.

Monte Carlo Expectation-Maximization (MCEM) Algorithm

When $p\left(\boldsymbol{u}_i \middle| 0, \boldsymbol{\Sigma}_u\right)$ or/and $p\left(\hat{\boldsymbol{\xi}}_i \middle| \boldsymbol{\xi}_i, \hat{\boldsymbol{B}}_{\xi i}, \hat{\boldsymbol{\Sigma}}_{\xi i}\right)$ does not follow a normal PDF, or even be unknown, the importance sampling can be used to approximate them, so that we can obtain the parameter point estimates using the Monte Carlo EM algorithm (Wei & Tanner, 1990). At the (*m*+1)th iteration of the MCEM algorithm, in the E-step, take the expectation of log-likelihood with respect to the posterior distribution of $\boldsymbol{u}_i$ and $\boldsymbol{\xi}_i$ as

$$E(\boldsymbol{\Psi} \mid \hat{\boldsymbol{\Psi}}^{(m)}) = \sum_{i=1}^{N} \int l(\boldsymbol{\Psi}, \boldsymbol{\xi}_i, \boldsymbol{u}_i)\, f(\boldsymbol{\xi}_i, \boldsymbol{u}_i \mid \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{B}}_{\xi i}, \hat{\boldsymbol{\Sigma}}_{\xi i}, \hat{\boldsymbol{\Psi}}^{(m)}) d\boldsymbol{u}_i d\boldsymbol{\xi}_i \text{ [10]}, \tag{13}$$

where $l(\boldsymbol{\Psi}, \boldsymbol{\xi}_i, \boldsymbol{u}_i)$ is the logarithm of the joint likelihood, and $f(\boldsymbol{\xi}_i, \boldsymbol{u}_i \mid \hat{\boldsymbol{\xi}}_i, \hat{B}_{\xi i}, \hat{\boldsymbol{\Sigma}}_{\xi i}, \hat{\boldsymbol{\Psi}}^{(m)})$ denotes the posterior distribution. Then we may draw *Q* samples of $\boldsymbol{\xi}_i^q$ and $\boldsymbol{u}_i^q$ (q = 1, …, *Q*) from a sampling distribution $H(\boldsymbol{\xi}_i, \boldsymbol{u}_i)$, which can be taken as the posterior or other distributions. The numerical approximation of Equation 13 can be rewritten as

$$E(\boldsymbol{\Psi} \mid \hat{\boldsymbol{\Psi}}^{(m)}) \approx \sum_{i=1}^{N} \left[\frac{1}{Q} \sum_{q=1}^{Q} l(\boldsymbol{\Psi}, \boldsymbol{\xi}_i^q, \boldsymbol{u}_i^q)\, f(\boldsymbol{\xi}_i^q, \boldsymbol{u}_i^q \mid \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{B}}_{\xi i}, \hat{\boldsymbol{\Sigma}}_{\xi i}, \hat{\boldsymbol{\Psi}}^{(m)}) \Big/ H(\boldsymbol{\xi}_i^q, \boldsymbol{u}_i^q) \right], \tag{14}$$

Especially, when $\boldsymbol{\xi}_i$ and $\boldsymbol{u}_i$ are both normally distributed, the integration in Equation 13 can be obtained easily when one samples directly from the posterior distribution, such that

$$E(\boldsymbol{\Psi} \mid \hat{\boldsymbol{\Psi}}^{(m)}) = \sum_{i=1}^{N} \left[\frac{1}{Q} \sum_{q=1}^{Q} l(\boldsymbol{\Psi}, \boldsymbol{\xi}_i^q, \boldsymbol{u}_i^q) \right], \tag{15}$$

[9] As there is no box-constraint can handle the non-negative matrix directly, in this article, we impose constraints on the variance and correlation terms.

[10] When random effects follow a (multivariate) normal distribution, we can obtain a closed form after integrate $\boldsymbol{u}$ in Equation 13 before approximating the expectation of log-likelihood. However, as the Monte Carlo approximation method (with 500 draws) is used to calculate Equation 14, the influence of integrating $\boldsymbol{u}$ firstly is negligible. Hence, in this article, both $\boldsymbol{\xi}_i$ and $\boldsymbol{u}_i$ are sampled from their posterior distributions. They can be drawn using the importance sampling when the normal assumption is violated.

where $(\boldsymbol{\xi}_i^q, \boldsymbol{u}_i^q)$ is the qth draw from the posterior distribution $f(\boldsymbol{\xi}_i, \boldsymbol{u}_i \mid \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{B}}_{\xi i}, \hat{\boldsymbol{\Sigma}}_{\xi i}, \hat{\boldsymbol{\Psi}}^{(m)})$ directly. Because $\boldsymbol{\xi}_i$ and $\boldsymbol{u}_i$ are from separate multivariate normal distributions, they can be drawn separately in the Monte Carlo sampling. That is, at the (m+1)th iteration, the posterior distribution of $\boldsymbol{\xi}_i$ is

$$L(\boldsymbol{\xi}_i \mid \hat{\boldsymbol{\Psi}}^{(m)}, \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{\Sigma}}_{\xi i}) \propto \phi(\boldsymbol{\xi}_i, \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(m)}, \boldsymbol{Z}\hat{\boldsymbol{\Sigma}}_u^{(m)}\boldsymbol{Z}^T + \hat{\sigma}^{2(m)}\boldsymbol{I}_T)\varphi(\hat{\boldsymbol{\xi}}_i, \boldsymbol{\xi}_i + \hat{\boldsymbol{B}}_{\xi i}, \hat{\boldsymbol{\Sigma}}_{\xi i}), \quad (16)$$

and the posterior distribution of $\boldsymbol{u}_i$ is

$$L(\boldsymbol{u}_i \mid \hat{\boldsymbol{\Psi}}^{(m)}, \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{\Sigma}}_{\xi i}) \propto \phi(\boldsymbol{\xi}_i, \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(m)} + \boldsymbol{Z}\boldsymbol{u}_i, \hat{\sigma}^{2(m)}\boldsymbol{I}_T)\phi(\boldsymbol{u}_i, \boldsymbol{0}, \hat{\boldsymbol{\Sigma}}_u^{(m)})\varphi(\hat{\boldsymbol{\xi}}_i, \boldsymbol{\xi}_i + \hat{\boldsymbol{B}}_{\xi i}, \hat{\boldsymbol{\Sigma}}_{\xi i}). \quad (17)$$

Given Equations 16 and 17, the Monte Carlo sampling continues as the following two sub steps:

(a) Draw $\boldsymbol{\xi}_i^q$ from a multivariate normal distribution with a mean vector $\left(\hat{\sigma}^{-2(m)}\boldsymbol{I}_T + \hat{\boldsymbol{\Sigma}}_{\xi i}^{-1}\right)^{-1}\left(\hat{\boldsymbol{\Sigma}}_{\xi i}^{-1}\left(\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{B}}_{\xi i}\right) + \hat{\sigma}^{-2(m)}\boldsymbol{I}_T(\boldsymbol{X}\hat{\boldsymbol{\beta}}^{(m)} + \boldsymbol{Z}\boldsymbol{u}_i^q)\right)$ and a covariance matrix $\left(\hat{\sigma}^{-2(m)}\boldsymbol{I}_T + \hat{\boldsymbol{\Sigma}}_{\xi i}^{-1}\right)^{-1}$.

(b) Draw $\boldsymbol{u}_i^q$ from a multivariate normal distribution with a mean vector $\left(\boldsymbol{Z}^T(\hat{\sigma}^{2(m)}\boldsymbol{I}_T)^{-1}\boldsymbol{Z} + (\hat{\boldsymbol{\Sigma}}_u^{(m)})^{-1}\right)^{-1}\left(\boldsymbol{Z}^T(\hat{\sigma}^{2(m)}\mathbf{I}_T)^{-1}(\boldsymbol{\xi}_i^q - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(m)})\right)$ and a covariance matrix $\left(\boldsymbol{Z}^T(\hat{\sigma}^{2(m)}\boldsymbol{I}_T)^{-1}\boldsymbol{Z} + (\hat{\boldsymbol{\Sigma}}_u^{(m)})^{-1}\right)^{-1}$.

In operation, the Monte Carlo sampling could be initiated by setting the first $\boldsymbol{\xi}_i^q$ as $\hat{\boldsymbol{\xi}}_i$, and drawing $\boldsymbol{u}_i^q$ conditioning on $\boldsymbol{\xi}_i^q$.

Under (multivariate) normal assumption, M-step proceeds with maximizing the conditional expectation in Equation 15 with respect to $\boldsymbol{\Psi}$. Given the form of $l(\boldsymbol{\Psi}, \boldsymbol{\xi}_i, \boldsymbol{u}_i)$, all the structural parameters (i.e., $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_u$ and σ^2) have the closed-form solutions as follows, which greatly simplifies the maximization step,

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \left(\sum_{q=1}^{Q} \sum_{i=1}^{N} \boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \times \sum_{q=1}^{Q} \sum_{i=1}^{N} \boldsymbol{X}^T \left[(\boldsymbol{\xi}_i^q)^T - \boldsymbol{Z}\boldsymbol{u}_i^q \right], \tag{18}$$

$$\hat{\sigma}^{2(m+1)} = \frac{\sum_{q=1}^{Q} \sum_{i=1}^{N} \left((\boldsymbol{\xi}_i^q)^T - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(m+1)} - \boldsymbol{Z}\boldsymbol{u}_i^q \right)^T \left((\boldsymbol{\xi}_i^q)^T - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(m+1)} - \boldsymbol{Z}\boldsymbol{u}_i^q \right)}{N \times Q \times T}, \tag{19}$$

$$\hat{\boldsymbol{\Sigma}}_u^{(m+1)} = \frac{\sum_{i=1}^{N} \sum_{q=1}^{Q} \boldsymbol{u}_i^q (\boldsymbol{u}_i^q)^T}{N \times Q}. \tag{20}$$

Wang et al.'s (2019) methods

As we mentioned before, Wang et al.'s method can be treated as a specific case of the proposed methods. Let $\boldsymbol{B} = \boldsymbol{0}$ in Equations 13, 16 and 17, the MMLE/MCEM algorithms will reduce to Wang et al.'s (2019) two-stage methods. Take Wang et al.'s (2019) MMLE method (denoted as "MMLE-U") as an example, the closed-form log marginal likelihood, which need to be maximized, is

$$\begin{aligned} & l_{MMLE-U}(\boldsymbol{\Psi}) = \log L_{MMLE-U}(\boldsymbol{\Psi}) \\ & \propto -\log \left| \sigma^2 \boldsymbol{I}_T + Z\boldsymbol{\Sigma}_u Z^T + \hat{\boldsymbol{\Sigma}}_{\xi i} \right| - (\hat{\boldsymbol{\xi}}_i - \boldsymbol{X}\boldsymbol{\beta})^T (\sigma^2 \boldsymbol{I}_T + Z\boldsymbol{\Sigma}_u Z^T + \hat{\boldsymbol{\Sigma}}_{\xi i})^{-1} (\hat{\boldsymbol{\xi}}_i - \boldsymbol{X}\boldsymbol{\beta}) \end{aligned}. \tag{21}$$

Naïve Method

If only individual point estimates of latent variables are the outcomes from stage I, the naïve method can be used to estimate $\boldsymbol{\Psi}$. Then $\hat{\boldsymbol{\beta}} = (\dot{\boldsymbol{X}}^T \dot{\boldsymbol{V}}^{-1} \dot{\boldsymbol{X}})^{-1} \dot{\boldsymbol{X}}^T \dot{\boldsymbol{V}}^{-1} \hat{\dot{\boldsymbol{\xi}}}$, which has been mentioned before. And the ML estimates of σ^2 and $\boldsymbol{\Sigma}_u$ are estimated by maximizing the following likelihood function,

$$F_{\text{naïve}} = -\log \left| \boldsymbol{Z}\boldsymbol{\Sigma}_u \boldsymbol{Z}^T + \sigma^2 \boldsymbol{I}_T \right| - \frac{1}{N} \operatorname{tr} \left\{ \left[\sum_i \left(\hat{\boldsymbol{\xi}}_i - \boldsymbol{X}\boldsymbol{\beta} \right)^T \left(\hat{\boldsymbol{\xi}}_i - \boldsymbol{X}\boldsymbol{\beta} \right) \right] \left(\boldsymbol{Z}\boldsymbol{\Sigma}_u \boldsymbol{Z}^T + \sigma^2 \boldsymbol{I}_T \right)^{-1} \right\},$$

where "tr" denotes the trace of a matrix.

Moment Estimation Method (MEM)

If population mean and covariance, instead of individual point estimates of latent variables, are estimated from raw response data in stage I, moment estimation method

(MEM) can be used in stage II, which is consistent with the traditional wisdom in SEM. The detailed EM algorithm to obtain population mean and covariance (i.e., $\boldsymbol{\mu}_{\xi}, \boldsymbol{\Sigma}_{\xi}$) is provided in Supplementary A, in which the population distribution is restricted to be multivariate normal. As a result, measurement biases and errors of individual point estimates can be considered to be plugged into population variables (i.e., mean and covariance).

Then the estimates of the population parameters of the overall abilities (i.e., $\hat{\boldsymbol{\mu}}_{\xi}$ and $\hat{\boldsymbol{\Sigma}}_{\xi}$) were input into $\hat{\boldsymbol{\mu}}_{\xi} = \boldsymbol{X}\boldsymbol{\beta}$ and $\hat{\boldsymbol{\Sigma}}_{\xi} = \boldsymbol{Z}\boldsymbol{\Sigma}_u\boldsymbol{Z}^T + \sigma^2\boldsymbol{I}_T$ to estimate the target structural model parameters. Hence, we can obtain $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\hat{\boldsymbol{\mu}}_{\xi}$, where $\boldsymbol{V}$ is defined in Equation 9, and the ML estimates of σ^2 and $\boldsymbol{\Sigma}_u$ through maximizing the following likelihood function $F_{MEM} = -\log\left|\boldsymbol{Z}\boldsymbol{\Sigma}_u\boldsymbol{Z}^T + \sigma^2\boldsymbol{I}_T\right| - \mathrm{tr}\left[\hat{\boldsymbol{\Sigma}}_{\xi}\left(\boldsymbol{Z}\boldsymbol{\Sigma}_u\boldsymbol{Z}^T + \sigma^2\boldsymbol{I}_T\right)^{-1}\right]$.

Both procedures above can be implemented using any off-the-shelf SEM packages. In this article, R package "lavaan" (Rosseel, 2012) was used to implement the naïve method and the MEM.

4. Simulation Study

A simulation study was conducted to evaluate the performances of the proposed two-stage methods, and compare them with the MCMC algorithm, the naïve two-stage method, the moment estimation method (MEM) as well as Wang et al.'s (2019) measurement error correction two-stage estimation method[11] (i.e., "MMLE-U" mentioned before). The detailed MCMC algorithm was presented in Appendix B. The naïve two-stage method, the MEM were implemented by R package "lavaan", the MMLE and MMLE-U methods were implemented by R package "stats", and the MCMC and MCEM methods were implemented

[11] In Wang et al. (2009), the performances of MMLE and MCEM were similar, but MCEM is more time consuming than MMLE. Hence, only MMLE, which corrected measurement errors solely, was compared in this article.

by Matlab. Without loss of generality, the latent growth curve (LGC) model was used as a special case of the LME models.

With the moment estimation method (MEM), we first computed the population parameters (i.e., the mean and variance-covariance matrix) of the overall abilities and residuals through Equation 3 by the EM algorithm, in which the population distribution is restricted to be multivariate normal. Then the estimates of the population parameters of the overall abilities (i.e., $\hat{\boldsymbol{\mu}}_{\xi}$ and $\hat{\boldsymbol{\Sigma}}_{\xi}$) were input into the stage II estimation process.

Simulation Design

The fixed and manipulated factors in the study were drawn from the previous literature. Three factors were manipulated: examinee sample size (200 *vs.* 2,000), test length (30 *vs.* 60), the regression coefficient between first-order and higher-order factors (i.e., four levels, $\lambda_k = \sqrt{0.8}, \sqrt{0.5}, \sqrt{0.3}$ and $U(0.55, 0.9)$[12]), and covariance matrix of $\boldsymbol{\Sigma}_u$ (Raudenbush & Liu, 2000; Ye, 2016),

$$\begin{bmatrix} 0.2 & 0.05 \\ 0.05 & 0.1 \end{bmatrix} \text{ (medium)}, \quad \begin{bmatrix} 0.1 & 0.025 \\ 0.025 & 0.05 \end{bmatrix} \text{ (small)}.$$

In terms of fixed effects, the mean intercept was set at 0 (i.e., β_0=0), and mean slope was set at 0.15 (i.e., β_1=0.15). Given the medium slope variance of 0.1 specified above, the mean slope of 0.15 leads to a medium standardized effect size of 0.5 (Raudenbush & Liu, 2000). Regarding the item parameters, *a*-parameters were drawn from $U(1.5, 2.5)$, *b*-parameters were drawn from $N(0, 1)$ (Cai, 2010), and *c*-parameters were drawn from $U(0.1, 0.2)$. Residual variance was $\sigma^2 = 0.15$ (Kohli et al., 2015). The number of measurement waves was fixed at 4 (Khoo, Wes, Wu, & Kwok, 2006; Ye, 2016), and test length was fixed at 60, with equal number of items loading on each of the three dimensions.

[12] The range of this uniform distribution depends on the first three values of λ, which can cover manipulated values of λ in this study.

Crossing all manipulated factors resulted in 32 conditions. 50 replications[13] were conducted for each condition. For all the conditions, the mean square error (MSE)[14] and bias of structural parameters were calculated to evaluate their performances. Take θ_s as an example of the parameter of interest, then the MSE and bias can be expressed respectively as

$$MSE_\theta = \frac{1}{n}\sum_{s=1}^{n}\left(\hat{\theta}_s - \theta_s\right)^2 \quad \text{and} \quad bias_\theta = \frac{1}{n}\sum_{s=1}^{n}\left(\hat{\theta}_s - \theta_s\right).$$

Simulation Results

Tables 1 - 4 provided the summary of structural parameter recovery in stage II under different conditions. Summary of latent trait and residual recovery in stage I was shown in the supplementary material.

Insert Tables 1-4 here.

In terms of MSEs, large sample size would lead to good performances of the MMLE, MMLE-U and naïve methods, and poor performance of the MCMC algorithm, the influence of sample size on other methods was negligible. Small covariance matrix can reduce the MSEs except for the MMLE and MMLE-U methods. When λ_k became smaller, the recoveries of $\boldsymbol{\beta}$ and τ_{00} became poorer, the recovery of σ^2 using the MCEM and naïve methods became poorer, and the recovery of σ^2 using the MMLE and MMLE-U methods became better. The biggest difference among different two-stage methods lies in the estimates of σ^2, because the estimation accuracy of variance-covariance matrix of $\boldsymbol{\xi}$ has less influence on $\boldsymbol{\Sigma}_u$ than σ^2.

Regarding biases, the MEM produces parameter estimates with smallest bias because this

[13] As the MCEM method is time-consuming to obtain estimators, 50 replications were done. Take the simplest condition as an example (i.e., $N = 200$, $J = 30$, $\lambda_k = \sqrt{0.8}$, medium covariance matrix), the cost of the MMLE and MCEM (with 500 draws) methods for one replication were 1.69 mins and 47.51 mins, respectively. More complex condition will lead to more computational time. On the other hand, a simulation check has been done for 500 replications using other computationally intensive methods, those methods performed similarly as those presented in this article (i.e., 50 replications). Overall, 50 replications were acceptable in this article.

[14] The mean square error (MSE) is used to evaluate the performances of those methods, as it contains the information of both bias and standard error (SE).

method does not rely on the assumption of normal measurement errors. The naïve method overestimates σ^2 under all the conditions unsurprisingly. Because the MAP estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ from HO-IRT contain relatively high bias (see Supplemental material) compared to $\boldsymbol{\theta}$ from UIRT or MIRT models (e.g., see Wang et al., 2019), without correcting bias in the stage II estimation, the parameter estimates from the MMLE-U method have higher bias than those from the MMLE method. In other words, correcting solely measurement error in stage II is inadequate when $\hat{\boldsymbol{\xi}}$ contains non-ignorable bias.

When test length is small (i.e., 30; Tables 3 - 4), the trend of parameter recovery was similar as the large test length case, and the influence of test length is negligible. Compared to the large test length case, the MCMC algorithm and the MEM performed similarly, the MMLE and MMLE-U methods performed poorer to estimate σ^2 and better to estimate τ_{00}, and other methods performed a little poorly with larger absolute values of MSEs and biases. The largest absolute values were mostly obtained by the MCMC algorithm or the MMLE/MMLE-U method. It appears that adding the correction of measurement bias into the two-stage method makes parameter estimation more consistent.

On the other hand, for the medium covariance matrix case with large test length (i.e., Table 1), under the conditions with small sample size, when $\lambda_k = \sqrt{0.5}$, in 10% of the replications, $\hat{\sigma}^2$ reached the lower bound (i.e., 0.001) from the MMLE method, in contrast to 8% from the MMLE-U method; the proportions of which $\hat{\sigma}^2$ reached the lower bound from both the MMLE and MMLE-U methods were 2%, 20%, and 4%, when $\lambda_k = \sqrt{0.8}$, $\sqrt{0.3}$ and $\lambda_k \sim U(0.55, 0.9)$, respectively. And under the conditions with medium sample size, when $\lambda_k = \sqrt{0.3}$, 2% $\hat{\sigma}^2$ reached the lower-bound by MMLE. Among other conditions, the MMLE/MMLE-U methods can recover σ^2 well in all replications. When an estimate reaches the boundary, this signals that the optimization is trapped in a local

optimum.

Similarly, across Tables 1 – 4, small λ_k, small sample size or small test length led to large proportion of which $\hat{\sigma}^2$ reached the lower bound using the MMLE/MMLE-U method. And the largest value was 34% obtained by the MMLE-U method when $\lambda_k = \sqrt{0.3}$ for small covariance matrix with small sample size and small test length. The influence of correcting measurement bias was negligible.

5. Real Data Analysis

This example is based on the National Educational Longitudinal Study 88 (NELS 88)[15], which was designed to track a nationally representative sample of approximately 24,500 students via multiple cognitive batteries from 8*th* to 12*th* grade (the first three studies) in years 1988, 1990, and 1992. Mathematics, Science, Reading and Social study were measured. Among them, as Mathematics and Reading had multiple levels in 1990 and 1992, according to the result of model-data fit analysis and item overlap rates, the medium level tests of Mathematics, the high level tests of Reading, Science and Social study were chosen to be analyzed in this article. The sample size was 410 after initial data cleaning, we used listwise deletion to eliminate the effect of missing data. The data contains binary responses to 116 items in each year, in which 21 items measured Reading, 25 items measured Science, 30 items measured Social study, and 40 items measured Mathematics sequentially. The true values of item parameters are from NELS 88 psychometrics report (https://nces.ed.gov/pubs95/95382.pdf). The higher-order IRT model was used on these cognitive batteries assuming the general factor represents overall academic achievement,

[15] In this example, even though a portion of the same items were used over time, since the measurement were given one year apart, the item level residual correlation is too weak to be considered. In addition, Wang, Kohli and Henn (2016) compared the model with vs. without nuisance factors using the same data set and showed that model without nuisance factors was preferred. Hence, in this section, the model without nuisance factors was focused on.

whereas the lower-order factors represent subject-specific proficiencies.

The item parameters from NELS 88 were calibrated by the UIRT model. Due to the between-item multidimensional nature of the HO-IRT model, the item parameters from UIRT and HO-IRT should be close even though the standard error of item parameters estimates from HO-IRT may be smaller due to borrowing strength from correlated domains. Hence, in this article, the item parameter values of the UIRT model from NELS psychometrics report were used as the "true" item parameter values of the HO-IRT model. The mean discrimination parameters were 0.974, 0.937, and 0.943 for the three measurement occasions, with their standard deviation of 0.326, 0.316, and 0.310. The mean and standard deviation of difficulty parameters were (0.575, 0.580, 0.834) and (0.850, 0.633, 0.863), respectively. The mean and standard deviation of guessing parameters were (0.185, 0.183, 0.184) and (0.133, 0.124, 0.107), respectively.

The MCMC algorithm was used in stage I to estimate the latent coefficients $\hat{\lambda}_{MCMC} = [0.941, 0.561, 0.924, 0.933]$ in the HO-IRT model, where the prior of $\boldsymbol{\lambda}$ was $U(0, 1)$. In stage I, the MAP method was used to estimate the person parameters (i.e., overall abilities and residuals), in which a standard normal distribution and $N(0, 1 - \lambda_k^2)$ ($k = 1, 2, 3, 4$) were used as the prior distributions of $\boldsymbol{\xi}$ and $\boldsymbol{\varepsilon}$, and the EM algorithm was used to estimate the population mean and covariance matrix of $\boldsymbol{\xi}$ and $\boldsymbol{\varepsilon}$.

Totally, six methods are compared in this section, they are the MCMC algorithm (on the integrative model), the EM algorithm in stage I with the MEM in stage II, and the MAP algorithm is stage I with MMLE, MCEM, MMLE-U, and naïve methods in stage II.

Insert Table 5 here.

Table 5 presented the mean and covariance matrix estimates of overall abilities in stage I from different methods, and Figures 1 - 2 provided the corresponding PDF plots from MCMC and MAP at different occasions. The means of $\hat{\xi}$ from the MCMC algorithm

increase over time, and the means from the MAP method and EM algorithm seem similar and no increasing trend. The sample variances of $\hat{\xi}$ calculated by the MAP method are the largest, and the population variances of $\hat{\xi}$ estimated by the EM algorithm is the smallest.

When to estimate parameters using MAP in stage I, the mean values of domain-specific abilities measured Reading are $[1.363, 0.992, 1.034]^T$, those measured Science are $[0.005, 0.281, 0.525]^T$, those measured Social study are $[1.531, 0.930, 1.052]^T$, and those measured Mathematics are $[0.783, 0.874, 1.051]^T$.

Insert Figures 1-2 here.

The point estimates of structural parameters of the NELS 88 data were shown in Table 6. As shown, $\hat{\boldsymbol{\beta}}$ s estimated from different methods are similar. Because of three large values of $\boldsymbol{\lambda}$, $\hat{\sigma}^2$ and $\hat{\tau}_{00}$ obtained by MMLE and MMLE-U are much larger than those obtained by other methods, which is consistent with simulation results. According to the simulation results, we know that large correlation between domain-specific ability and overall ability may lead to inaccurate estimations of residuals and variances of random effects using the MEM. The negative estimations of random effects' variances from the MEM are possible, because the multivariate normal assumption in stage I may not be satisfied in this illustration (see Figures 1 - 2).

Insert Table 6 here.

6. Conclusions

In this paper, we extend Wang et al. (2019)'s study by proposing two new model estimation methods for data analysis when the outcome variable in a linear mixed effects model is latent and therefore measured with error. Both of them fall within the framework of two-stage estimation, and they are based on the maximum likelihood estimation scheme. Compare to the two-stage MMLE or MCEM method proposed in Wang et al. (2019) that

only takes into account measurement error of the dependent variable, the new methods handle both measurement bias and errors in stage II estimation. Therefore, our new methods will be especially preferable when the latent variable (i.e., latent trait $\boldsymbol{\theta}$ or $\boldsymbol{\xi}$) from stage I contains non-ignorable bias, which happens when test length is short. Because the overall ability and domain-specific ability are linearly related, the domain-specific ability growth can be obtained based on the following equation: $\boldsymbol{\theta_i} = \boldsymbol{\lambda}(\boldsymbol{X\beta} + \boldsymbol{Zu_i} + \boldsymbol{e}_i) + \boldsymbol{\varepsilon}$. In other words, the proposed methods can also be applied to obtain the domain-specific ability growth easily.

Across all simulation conditions, if the measurement biases and errors follow a multivariate normal distribution, and there is no additional covariate in structural models, the moment estimation method is recommended; otherwise, the Monte Carlo EM algorithm is recommended to correct measurement bias and error simultaneously.

Similar to Wang, et al. (2019), the MMLE method builds upon the (multivariate) normal assumption such that the marginal likelihood has a closed form. When measurement biases and errors or/and random effects do not follow a (multivariate) normal distribution, the importance sampling can be used to numerically approximate the integration in the Monte Carlo EM algorithm. Comparing to the MMLE method, the MCEM method has a greater flexibility because of avoiding the (multivariate) normal assumption. On the other hand, as the MMLE estimators are obtained by maximizing the closed form of the marginal likelihood, they do not depend on the choice of the quadrature nodes, and the influence of the multidimensional structure of latent variables on the estimate process can be negligible. Therefore, the MMLE method outperforms the MCEM algorithm in terms of calculation time and efficiency.

This study can be extended in a number of directions. First, as the (multivariate) weighted maximum likelihood (WLE; Warm, 1989; Wang, 2015) produces an unbiased estimator, it can replace MAP in stage I such that we only need to correct measurement error

in stage II using Wang et al.'s methods (2019). Second, we assume item parameters are known and free of error, so that the influence of item parameter calibration errors vanishes. In case the calibration sample size is small, then the sampling error cannot be ignored, and we can either use the Bootstrap-calibrated interval estimates for the overall abilities and residuals (Liu & Yang, 2018) in stage I; or estimate the standard errors of item parameters using a multiple imputation based approach (Huang & Cai, 2019), and then incorporate the calibration standard error in the correction method before proceeding with the stage II estimation. Last, a simpler model with independent residuals (i.e., independent residuals of both e_{it} and ε_{ikt} over time) was considered in this article, this assumption needs to be checked in future studies by comparing models using AIC or BIC, as allowing residuals to correlate over time in longitudinal models is more flexible. Moreover, if the same items are used repeatedly and the adjacent measurements are closer in time, there may be residual correlations among responses on the same item over time (e.g., Cai, 2010; Wang, Kohli, & Henn, 2016). In this case, additional nuisance factors need to be introduced to account for such correlation.

Reference

Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of educational and behavioral Statistics*, 22(1), 47-76. doi: 10.3102/10769986022001047

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155-173. doi: 10.1007/BF02294170

Bacharach, V. R., Baumeister, A. A., & Furr, R. M. (2003). Racial and gender science achievement gaps in secondary education. *The Journal of Genetic Psychology*, *164*(1), 115–126. doi: 10.1080/00221320309597507

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, 28, 97-104.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61*, 109–121. doi: 10.1007/BF02296961

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological methods & research*, 5(1), 3-52. doi: 10.1177/004912417600500101

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, 75, 581-612.

Cai, L. (2010). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, *35*(3), 307-335. doi: 10.3102/1076998609353115

Carroll, R. J., Ruppert, D., & Stefanski, L. A. (1995). *Measurement error in nonlinear models: A modern perspective*. London: Chapman & Hall.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural equation modeling*, 12(3), 471-492. doi: 10.1207/s15328007sem1203_7

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189-225. doi: 10.1207/s15327906mbr4102_5

Cohen, A. S., Bottge, B. A., &Wells, C. S. (2001). Using item response theory to assess effects of mathematics instruction in special populations. *Exceptional Children*, *68*(1), 23–44. doi: 10.1177/001440290106800102.

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure modeling* (pp. 195–223). Mahwah, NJ: Erlbaum. Devlieger, I., Mayer, A., & Rosseel,

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer Science & Business Media.

de la Torre, J., & Song, H. (2009). Simultaneous estimation of overall and domain abilities: A higher-order IRT model approach. *Applied Psychological Measurement*, *33*(8), 620-639. doi: 10.1177/0146621608326423

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of personality and social psychology*, *91*(6), 1138. doi: 10.1037/0022-3514.91.6.1138

Fox, J. P., & Glas, C. A. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*(2), 271-288. doi: 10.1007/BF02294839

Fox, J. P., & Glas, C. A. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika*, *68*(2), 169-191. doi: 10.1007/BF02294796

Fuller, W. A. (1987). *Measurement error models*. New York: Wiley.

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, *32*(3), 505–523. doi: 10.2307/146181

Gotay, C. C., Blaine, D., Haynes, S. N., Holup, J., & Pagano, I. S. (2002). Assessment of quality of life in a multicultural cancer patient population. *Psychological Assessment*, *14*(4), 439. doi: 10.1037/1040-3590.14.4.439

Golay, P., & Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the French Wechsler Adult Intelligence Scale (WAIS-III). *Psychological assessment*, *23*(1), 143. doi: 10.1037/a0021230

Hills, P., & Argyle, M. (2002). The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being. *Personality and individual differences*, *33*(7), 1073-1082. doi: 10.1016/S0191-8869(01)00213-6

Huang, H. Y. (2015). A multilevel higher order item response theory model for measuring latent growth in longitudinal data. *Applied Psychological Measurement*, 39, 362-372. doi: 10.1177/0146621614568112

Huang, S., & Cai, L. (2019). Improving Standard Error Estimates in Multistage Estimation: A Multiple Imputation (MI) Based Approach. *Multivariate behavioral research*, *54*(1), 154-154. doi: 10.1080/00273171.2018.1557034

Huang, H. Y., Chen, P. H., & Wang, W. C. (2012). Computerized adaptive testing using a class of high-order item response theory models. *Applied Psychological Measurement*, *36*(8), 689-706. doi: 10.1177/0146621612459552

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 38(1), 79-93. doi: 10.1111/j.1745-3984.2001.tb01117.x

Khoo, S., West, S., Wu, W., & Kwok, O. (2006). Longitudinal methods. In M. Eid & E.

Diener (Eds.), *Handbook of psychological measurement: A multimethod perspective* (pp. 301–317). Washington, DC: APA.

Kohli, N., Hughes, J.,Wang, C., Zopluoglu, C., & Davison, M. L. (2015). Fitting a linear–linear piecewise growth mixture model with unknown knots: A comparison of two common approaches to inference. *Psychological Methods*, 20(2), 259. doi: 10.1037/met0000034

Lee, S. Y., & Song, X. Y. (2003). Bayesian analysis of structural equation models with dichotomous variables. *Statistics in medicine*, 22(19), 3073-3088. doi: 10.1002/sim.1544

Liu, Y., & Yang, J. (2018). Bootstrap-calibrated interval estimates for latent variable scores in item response theory. *Psychometrika*, 83, 333–354. doi: 10.1007/s11336-017-9582-9

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lu, I. R., & Thomas, D. R. (2008). Avoiding and correcting bias in score-based latent variable regression with discrete manifest items. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 462-490. doi: 10.1080/10705510802154323

Lu, I. R., Thomas, D. R., & Zumbo, B. D. (2005). Embedding IRT in structural equation models: A comparison with regression based on IRT scores. *Structural Equation Modeling*, *12*(2), 263-277. doi: 10.1207/s15328007sem1202_5

Mislevy, R. J. (1987). Recent developments in item response theory with implications for teacher certification. *Review of research in education*, 239-275. doi: 10.3102/0091732X014001239

Mislevy, R. J., & Sheehan, K. M. (1989). Information matrices in latent-variable models. *Journal of Educational and Behavioral Statistics, 14*(4), 335-350. doi: 10.3102/10769986014004335

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor

versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407-422. doi: 10.1016/j.intell.2013.06.004

Nussbaum, E. Michael; Hamilton, Laura S.; Snow, Richard E. Enhancing the validity and usefulness of large-scale educational assessments: IV NELS: 88 science achievement to 12th grade. *American Educational Research Journal*, 1997, 34.1: 151-173. doi: 10.3102/00028312034001151

Pastor, D. A. (2003). The use of multilevel item response theory modeling in applied research: An illustration. *Applied measurement in education*, 16(3), 223-243. doi: 10.1207/S15324818AME1603_4

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1), 12-35. doi: 10.1080/10618600.1995.10474663

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stat*a. New York: STATA Press.

Raudenbush, S.W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, 5(2), 199. doi: 10.1037/1082-989X.5.2.199

Raudenbush, S.W., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage. 2nd.

Rijmen, F., Vansteelandt, K., & De Boeck, P. (2008). Latent class models for diary method data: Parameter estimation by local computations. *Psychometrika*, 73(2), 167. doi: 10.1007/s11336-007-9001-8

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563-575. doi: 10.1007/BF02296196

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC.

Tian, W., Cai, L., Thissen, D., & Xin, T. (2013). Numerical differentiation methods for computing error covariance matrices in item response theory modeling: An evaluation and a new proposal. *Educational and Psychological Measurement*, *73*(3), 412-439. doi: 0.1177/0013164412465875

Tucker, L. (1971). Relations of factor score estimates to their use. *Psychometrika, 36,* 427–436. doi: 10.1007/BF02291367

Von Davier, M., & Sinharay, S. (2007). An importance sampling EM algorithm for latent regression models. *Journal of Educational and Behavioral Statistics*, 32(3), 233-251. doi: 10.3102/1076998607300422

Wang, C. (2015). On latent trait estimation in multidimensional compensatory item response models. *Psychometrika*, *80*(2), 428-449. doi: 10.1007/s11336-013-9399-0

Wang, C., Kohli, N., & Henn, L. (2016). A second-order longitudinal model for binary outcomes: Item response theory versus structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 455-465.

Wang, C., & Nydick, S. W. (2019). On Longitudinal Item Response Theory Models: A Didactic. *Journal of Educational and Behavioral Statistics*, doi: 10.3102/1076998619882026.

Wang, C., Xu, G. & Zhang, X. (2019). Correction for item response theory latent trait measurement error in linear mixed effects models. *Psychometrika*. 84: 673. doi: 10.1007/s11336-019-09672-7

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427-450. doi: 10.1007/BF02294627

Wei, G. C., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and

the poor man's data augmentation algorithms. *Journal of the American statistical Association*, *85*(411), 699-704. doi: 10.1080/01621459.1990.10474930

Ye, F. (2016). Latent growth curve analysis with dichotomous items: Comparing four approaches. *British Journal of Mathematical and Statistical Psychology*, 69, 43–61. doi: 10.1111/bmsp.12058

Tables and Figures

Table 1. Summary of parameter recovery for medium covariance matrix and large test length.

λ	Par.	MCMC		MEM		MMLE		MCEM		MMLE-U		Naïve	
		MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias
		N = 200											
$\sqrt{0.8}$	β_0	**.002**	.033	**.002**	**-.018**	.003	.024	.003	.027	.003	.031	.003	.030
	β_1	.005	-.069	**.001**	**-.012**	**.001**	-.021	**.001**	-.020	**.001**	-.026	**.001**	-.025
	σ^2	.009	-.091	.006	-.072	.029	.170	**.003**	-.051	.027	.163	**.003**	**.050**
	τ_{00}	.014	-.118	.005	-.061	.020	.125	**.003**	-.043	.022	.134	**.003**	**-.042**
	τ_{01}	.002	-.043	**.000**	**.005**	.001	.020	**.000**	-.016	.001	.016	.001	-.018
	τ_{11}	.003	-.055	**.001**	-.031	.002	.037	**.001**	-.035	**.001**	**.029**	.002	-.039
$\sqrt{0.5}$	β_0	**.002**	.038	.005	**-.031**	**.002**	.032	.003	.035	.003	.040	.003	.039
	β_1	.006	-.077	**.002**	**-.035**	.003	-.047	.003	-.046	.004	-.053	.003	-.052
	σ^2	.009	-.092	**.005**	**-.068**	.011	.099	.014	-.117	.008	.088	.013	.111
	τ_{00}	.016	-.128	.015	-.119	**.004**	-.016	.015	-.120	**.004**	**-.011**	.012	-.106
	τ_{01}	.003	-.051	**.000**	-.005	.001	**-.002**	.001	-.020	.001	-.005	.001	-.027
	τ_{11}	.004	-.061	.004	-.061	**.001**	**-.012**	.003	-.058	.001	-.019	.004	-.060
$\sqrt{0.3}$	β_0	.003	.042	.018	-.094	**.002**	**.025**	**.002**	.028	.003	.035	**.002**	.034
	β_1	.009	-.091	**.004**	**-.027**	.005	-.069	.005	-.067	.006	-.077	.006	-.076
	σ^2	.008	-.090	**.001**	**-.017**	**.001**	**-.017**	.017	-.128	.002	-.039	.013	.114
	τ_{00}	.018	-.135	.031	-.172	.010	-.084	.025	-.159	**.009**	**-.081**	.023	-.149
	τ_{01}	.003	-.059	**.001**	-.028	**.001**	**-.023**	**.001**	-.036	.001	-.026	.002	-.038
	τ_{11}	.004	-.067	.007	-.075	**.002**	**-.043**	.007	-.081	.002	-.047	.006	-.077
$U(0.55,0.9)$	β_0	**.003**	.043	.005	-.045	**.003**	**.028**	**.003**	.032	**.003**	.036	**.003**	.035
	β_1	.006	-.076	**.002**	**-.035**	**.002**	-.043	**.002**	-.042	.003	-.049	.003	-.049
	σ^2	.009	-.092	**.005**	**-.068**	.013	.108	.011	-.104	.011	.097	.010	.096
	τ_{00}	.016	-.127	.017	-.124	**.007**	**.011**	.012	-.103	.007	.017	.010	-.093
	τ_{01}	.003	-.051	**.000**	**-.006**	.001	**.006**	**.000**	-.018	.001	.002	.001	-.024
	τ_{11}	.004	-.061	.004	-.057	**.001**	**-.004**	.003	-.053	**.001**	-.010	**.003**	-.056
		N = 2,000											
$\sqrt{0.8}$	β_0	**.001**	.032	**.001**	-.018	**.001**	**.017**	**.001**	.019	**.001**	.024	**.001**	.024
	β_1	.004	-.065	**.000**	**-.012**	.001	-.023	.001	-.024	.001	-.028	.001	-.027
	σ^2	.011	-.102	**.000**	-.071	.028	.167	.003	-.052	.025	.159	.002	**.050**
	τ_{00}	.019	-.137	.003	-.060	.017	.129	.002	-.038	.021	.143	**.001**	**-.036**
	τ_{01}	.001	-.036	**.000**	**.006**	.001	.024	**.000**	-.015	**.000**	.019	**.000**	-.017
	τ_{11}	.004	-.065	**.001**	-.031	.002	.038	**.001**	-.033	.001	**.030**	**.001**	-.038
$\sqrt{0.5}$	β_0	.002	.036	.002	-.031	**.001**	**.019**	**.001**	.022	**.001**	.027	**.001**	.026
	β_1	.006	-.078	**.001**	**-.035**	.002	-.045	.002	-.045	.003	-.052	.003	-.051
	σ^2	.013	-.113	**.001**	**-.068**	.008	.091	.015	-.123	.006	.079	.010	.102
	τ_{00}	.023	-.150	.014	-.119	**.001**	-.005	.014	-.117	**.001**	**.000**	.010	-.101
	τ_{01}	.002	-.042	**.000**	-.005	**.000**	**-.004**	**.000**	-.018	**.000**	-.007	.001	-.029
	τ_{11}	.005	-.074	.004	-.061	**.000**	**-.012**	.004	-.060	**.000**	-.018	.004	-.060
$\sqrt{0.3}$	β_0	.002	.036	.008	-.094	**.001**	**.020**	**.001**	.022	.001	.031	**.001**	.029
	β_1	.008	-.090	**.003**	**-.037**	.005	-.070	.005	-.069	.006	-.078	.006	-.077
	σ^2	.015	-.120	.011	**-.016**	**.000**	-.018	.018	-.134	.002	-.040	.012	.111
	τ_{00}	.027	-.163	.027	-.172	.007	-.080	.027	-.163	**.006**	**-.077**	.021	-.145
	τ_{01}	.002	-.048	**.001**	-.028	**.001**	**-.023**	**.001**	-.035	**.001**	-.026	.002	-.039
	τ_{11}	.007	-.080	.007	-.085	**.002**	**-.045**	.007	-.081	.002	-.049	.006	-.077
$U(0.55,0.9)$	β_0	.001	.032	.003	-.045	**.000**	**.017**	**.001**	.020	.001	.025	.001	.023
	β_1	.006	-.075	**.001**	**-.035**	.002	-.041	.002	-.042	.002	-.048	.002	-.046
	σ^2	.013	-.112	**.001**	**-.068**	.013	.108	.012	-.120	.011	.097	.008	.091
	τ_{00}	.022	-.149	.012	-.124	**.003**	**.028**	.010	-.095	.003	.034	.008	-.084
	τ_{01}	.002	-.040	**.000**	-.006	**.000**	**.000**	**.000**	-.019	**.000**	-.003	.001	-.027
	τ_{11}	.005	-.073	.003	-.057	**.000**	**.001**	.003	-.052	**.000**	-.006	.003	-.054

Note. MCM denotes the moment estimation method. MMLE-U denotes the marginalized MLE when correcting measurement error solely. The bold value means the smallest value in a specific condition.

Table 2. Summary of parameter recovery for small covariance matrix with large test length.

λ	Par.	MCMC		MEM		MMLE		MCEM		MMLE-U		Naïve	
		MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias
		N = 200											
$\sqrt{0.8}$	β_0	**.002**	.033	**.002**	**-.016**	**.002**	.017	**.002**	.020	**.002**	.023	**.002**	.022
	β_1	.005	-.069	**.001**	**-.012**	**.001**	-.019	**.001**	-.021	**.001**	-.024	**.001**	-.022
	σ^2	.009	-.094	.006	-.075	.056	.236	**.003**	-.053	.054	.230	**.003**	**.051**
	τ_{00}	.002	-.043	.002	.043	.004	-.012	**.001**	-.017	.004	**-.008**	**.001**	-.021
	τ_{01}	.001	-.028	.001	.029	.001	-.009	**.000**	-.011	.001	-.010	**.000**	**-.008**
	τ_{11}	**.000**	-.020	.001	.020	.001	**-.008**	**.000**	-.011	.001	-.012	**.000**	-.013
$\sqrt{0.5}$	β_0	.002	.030	.004	-.026	**.001**	**.018**	**.001**	.020	.002	.026	.002	.025
	β_1	.006	-.076	**.002**	**-.033**	.003	-.048	.003	-.048	.003	-.054	.003	-.053
	σ^2	.008	-.091	**.005**	**-.066**	.026	.157	.014	-.119	.022	.145	.012	.105
	τ_{00}	**.002**	-.042	**.002**	-.031	**.002**	**.013**	.003	-.057	**.002**	.015	.003	-.051
	τ_{01}	.001	-.036	.001	.018	.001	**-.004**	**.000**	-.013	.001	-.005	**.000**	-.016
	τ_{11}	.001	-.021	**.000**	-.013	**.000**	.009	.001	-.026	**.000**	**.006**	.001	-.025
$\sqrt{0.3}$	β_0	.002	.035	.011	-.064	.002	**.018**	**.001**	.019	.002	.028	.002	.027
	β_1	.008	-.091	**.004**	**-.050**	.005	-.069	.005	-.069	.006	-.076	.006	-.076
	σ^2	.009	-.092	**.001**	**-.014**	.002	.040	.017	-.131	**.001**	.017	.013	.115
	τ_{00}	**.002**	-.043	.005	-.061	.003	**-.020**	.006	-.079	**.002**	**-.020**	.006	-.072
	τ_{01}	.002	-.038	**.000**	**-.006**	.001	-.016	.001	-.024	.001	-.016	**.000**	-.019
	τ_{11}	.001	-.022	.001	-.035	**.000**	**-.015**	.002	-.041	**.000**	-.018	**.001**	-.038
$U(0.55,0.9)$	β_0	**.002**	.030	.003	-.027	**.002**	**.019**	**.002**	.023	**.002**	.027	**.002**	.025
	β_1	.006	-.073	**.002**	**-.033**	**.002**	-.037	**.002**	-.038	**.002**	-.042	**.002**	-.041
	σ^2	.009	-.096	**.007**	**-.077**	.037	.190	.011	-.102	.034	.180	.009	.093
	τ_{00}	.002	-.045	**.001**	**-.016**	.006	.041	.002	-.034	.006	.043	.002	-.039
	τ_{01}	.001	-.033	.001	.021	.001	.008	**.000**	-.012	.001	**.006**	**.000**	-.011
	τ_{11}	**.000**	-.020	**.000**	**-.002**	.001	.008	.001	-.025	**.000**	.005	.001	-.026
		N = 2,000											
$\sqrt{0.8}$	β_0	.001	.029	.001	-.020	**.000**	**.013**	**.000**	.015	.001	.020	**.000**	.019
	β_1	.004	-.064	**.000**	**-.015**	**.000**	-.018	.000	-.020	.001	-.022	**.000**	-.021
	σ^2	.011	-.104	**.000**	**-.015**	.056	.237	.003	-.050	.054	.231	.003	.053
	τ_{00}	.004	-.064	.002	.045	.001	-.009	**.000**	-.018	**.000**	**-.005**	**.000**	-.019
	τ_{01}	**.000**	-.020	.001	.028	**.000**	**-.007**	**.000**	-.008	**.000**	-.009	**.000**	-.008
	τ_{11}	.001	-.031	**.000**	.018	**.000**	**-.013**	**.000**	**-.013**	**.000**	-.017	**.000**	-.015
$\sqrt{0.5}$	β_0	.002	.038	.003	-.045	**.000**	**.016**	.001	.017	.001	.024	**.001**	.023
	β_1	.006	-.079	**.001**	**-.030**	.002	-.042	.002	-.043	.002	-.048	.002	-.047
	σ^2	.013	-.116	**.001**	**.026**	.026	.161	.016	-.125	.022	.148	.012	.107
	τ_{00}	.005	-.069	**.001**	-.017	**.001**	**.012**	.003	-.054	**.001**	.015	.002	-.049
	τ_{01}	.001	-.024	**.000**	.011	**.000**	.002	**.000**	-.011	**.000**	**.001**	**.000**	-.014
	τ_{11}	.001	-.034	**.000**	-.010	**.000**	.002	.001	-.028	**.000**	**-.000**	.001	-.028
$\sqrt{0.3}$	β_0	.002	.035	.008	-.078	**.000**	**.015**	.001	.017	.001	.026	**.001**	.025
	β_1	.008	-.091	**.003**	**-.050**	.004	-.066	.005	-.066	.005	-.074	.005	-.073
	σ^2	.015	-.121	.010	.096	.002	.047	.018	-.134	**.001**	**.023**	.013	.114
	τ_{00}	.005	-.073	.003	-.057	**.002**	-.035	.007	-.081	**.002**	**-.034**	.005	-.073
	τ_{01}	.001	-.028	**.000**	-.010	**.000**	**-.009**	.001	-.023	**.000**	-.011	**.000**	-.019
	τ_{11}	.001	-.037	.001	-.031	**.000**	**-.020**	.002	-.042	**.000**	-.022	.001	-.038
$U(0.55,0.9)$	β_0	.002	.037	.003	-.044	**.000**	**.014**	.001	.016	.001	.023	.001	.021
	β_1	.006	-.077	**.001**	**-.032**	.002	-.039	.002	-.040	.002	-.045	.002	-.043
	σ^2	.013	-.113	**.001**	**.015**	.033	.179	.013	-.111	.029	.167	.010	.099
	τ_{00}	.005	-.067	**.001**	**-.008**	.002	.026	.002	-.045	.002	.028	.002	-.044
	τ_{01}	.001	-.024	**.000**	.016	**.000**	.005	**.000**	-.012	**.000**	**.003**	**.000**	-.012
	τ_{11}	.001	-.034	**.000**	**-.005**	**.000**	.008	.001	-.025	**.000**	**.005**	.001	-.026

Note. MCM denotes the moment estimation method. MMLE-U denotes the marginalized MLE when correcting measurement error solely.

Table 3. Summary of parameter recovery for medium covariance matrix and small test length.

λ	Par.	MCMC		MEM		MMLE		MCEM		MMLE-U		Naïve	
		MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias
		N = 200											
$\sqrt{0.8}$	β_0	**.002**	.025	**.002**	**-.014**	**.002**	.024	.003	.026	.003	.036	.003	.035
	β_1	.004	-.061	**.001**	**-.016**	.002	-.032	.002	-.033	.002	-.039	.002	-.038
	σ^2	.008	-.090	**.003**	**-.047**	.027	.162	.005	-.066	.022	.148	**.003**	.054
	τ_{00}	.014	-.119	.005	-.060	.016	.104	.004	-.051	.019	.119	**.003**	**-.048**
	τ_{01}	.002	-.043	**.000**	**.004**	.001	.015	.001	-.018	.001	.009	.001	-.022
	τ_{11}	.003	-.053	.001	-.029	**.001**	.024	.002	-.040	**.001**	**.013**	.002	-.047
$\sqrt{0.5}$	β_0	**.002**	**.030**	.005	-.035	.004	.037	.004	.041	.004	.050	.004	.047
	β_1	.006	-.075	**.002**	**-.028**	.004	-.055	.004	-.054	.005	-.064	.004	-.062
	σ^2	.008	-.088	**.004**	**-.049**	.008	.084	.015	-.119	.005	.062	.009	.095
	τ_{00}	.017	-.128	.015	-.118	.005	-.018	.015	-.120	**.004**	**-.016**	.014	-.114
	τ_{01}	.003	-.054	**.000**	**-.005**	.001	-.011	.001	-.023	.001	-.014	.001	-.033
	τ_{11}	.004	-.060	.004	-.059	**.001**	**-.022**	.005	-.066	**.001**	-.029	.005	-.067
$\sqrt{0.3}$	β_0	.003	.038	.010	-.049	**.002**	**.027**	**.002**	.029	.003	.045	.003	.044
	β_1	.009	-.089	**.003**	**-.033**	.005	-.066	.005	-.066	.007	-.079	.007	-.079
	σ^2	.007	-.084	.002	**-.010**	**.001**	-.016	.017	-.130	.004	-.058	.010	.097
	τ_{00}	.017	-.131	.028	-.160	.012	-.094	.027	-.164	**.011**	**-.091**	.025	-.156
	τ_{01}	.004	-.060	**.001**	**-.026**	**.001**	**-.026**	.002	-.038	**.001**	-.029	.002	-.040
	τ_{11}	.004	-.063	.007	-.083	**.003**	**-.051**	.007	-.082	.004	-.057	.007	-.083
U(0.55,0.9)	β_0	**.003**	**.028**	.005	-.037	**.003**	**.028**	**.003**	.031	.004	.043	**.003**	.042
	β_1	.006	-.074	**.002**	**-.026**	**.002**	-.039	**.002**	-.039	.003	-.050	.003	-.049
	σ^2	.008	-.088	**.003**	**-.047**	.014	.110	.012	-.106	.010	.089	.008	.089
	τ_{00}	.016	-.127	.014	-.115	**.005**	**-.001**	.013	-.107	**.005**	.005	.012	-.102
	τ_{01}	.003	-.050	**.000**	**-.002**	.001	-.005	.001	-.024	.001	-.010	.001	-.031
	τ_{11}	.004	-.059	.003	-.052	**.001**	**-.012**	.004	-.059	.001	-.020	.004	-.063
		N = 2,000											
$\sqrt{0.8}$	β_0	**.001**	.025	**.001**	**-.017**	**.001**	.027	.002	.030	.002	.039	.002	.037
	β_1	.004	-.064	**.000**	**-.013**	.001	-.028	.001	-.030	.001	-.037	.001	-.035
	σ^2	.011	-.103	**.000**	**.001**	.026	.162	.004	-.065	.023	.150	.003	.057
	τ_{00}	.018	-.135	**.003**	-.054	.012	.102	**.003**	-.052	.014	.113	**.003**	**-.050**
	τ_{01}	.001	-.036	**.000**	**.003**	**.000**	.015	**.000**	-.018	**.000**	.010	**.000**	-.022
	τ_{11}	.005	-.067	.001	-.029	.001	.023	.002	-.039	**.000**	**.013**	.002	-.047
$\sqrt{0.5}$	β_0	**.002**	**.034**	.003	-.041	**.002**	.037	**.002**	.040	.003	.051	.003	.048
	β_1	.007	-.079	**.001**	**-.025**	.003	-.049	.003	-.049	.004	-.059	.004	-.058
	σ^2	.013	-.113	**.002**	**.043**	.007	.082	.016	-.125	.004	.059	.009	.097
	τ_{00}	.023	-.151	.014	-.117	**.001**	-.018	.016	-.125	**.001**	**-.014**	.012	-.111
	τ_{01}	.002	-.042	**.000**	-.013	**.000**	**-.010**	.001	-.021	**.000**	-.014	.001	-.033
	τ_{11}	.006	-.075	.004	-.061	**.001**	**-.021**	.004	-.065	.001	-.027	.004	-.066
$\sqrt{0.3}$	β_0	**.001**	**.032**	.010	-.080	.002	.039	.002	.041	.004	.055	.003	.052
	β_1	.008	-.091	**.003**	**-.040**	.006	-.076	.006	-.076	.008	-.088	.008	-.087
	σ^2	.014	-.118	.016	.126	**.001**	**-.022**	.017	-.131	.004	-.063	.009	.094
	τ_{00}	.026	-.162	.027	-.162	**.008**	**-.085**	.026	-.161	**.008**	-.086	.023	-.152
	τ_{01}	.002	-.049	**.001**	-.033	**.001**	**-.027**	**.001**	-.038	**.001**	-.030	.002	-.041
	τ_{11}	.007	-.081	.007	-.082	**.003**	**-.051**	.007	-.083	**.003**	-.056	.007	-.082
U(0.55,0.9)	β_0	**.001**	**.028**	.004	-.048	**.001**	.030	.002	.032	.002	.045	.002	.044
	β_1	.005	-.073	**.001**	**-.025**	.002	-.045	.002	-.045	.003	-.055	.003	-.054
	σ^2	.013	-.112	**.002**	**.041**	.010	.096	.014	-.117	.007	.075	.008	.089
	τ_{00}	.022	-.148	.012	-.107	**.003**	**.014**	.012	-.105	**.003**	.017	.010	-.097
	τ_{01}	.002	-.042	**.000**	-.012	**.000**	**-.010**	.001	-.024	**.000**	-.014	.001	-.033
	τ_{11}	.005	-.073	.003	-.057	**.001**	**-.014**	.004	-.060	.001	-.022	.004	-.063

Note. MCM denotes the moment estimation method. MMLE-U denotes the marginalized MLE when correcting measurement error solely.

Table 4. Summary of parameter recovery for small covariance matrix with small test length.

λ	Par.	MCMC		MEM		MMLE		MCEM		MMLE-U		Naïve	
		MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias
		N = 200											
$\sqrt{0.8}$	β_0	**.001**	.025	.002	-.019	**.001**	**.014**	.002	.017	.002	.027	.002	.025
	β_1	.004	-.062	**.001**	**-.012**	**.001**	-.022	**.001**	-.025	**.001**	-.030	**.001**	-.028
	σ^2	.009	-.093	**.001**	**-.023**	.055	.233	.004	-.064	.049	.221	.004	.061
	τ_{00}	.002	-.042	.002	-.035	.006	-.026	**.001**	**-.021**	.005	-.023	.002	-.026
	τ_{01}	.001	-.029	**.000**	**.001**	.001	-.014	**.000**	-.012	.001	-.016	**.000**	-.010
	τ_{11}	**.000**	-.019	**.000**	-.018	.001	-.019	**.000**	**-.014**	.001	-.023	**.000**	-.018
$\sqrt{0.5}$	β_0	**.002**	.029	.004	**-.027**	.004	.034	.004	.038	.004	.048	.004	.045
	β_1	.006	-.075	**.002**	**-.035**	.003	-.049	.003	-.050	.004	-.059	.004	-.057
	σ^2	.008	-.091	**.001**	**-.022**	.024	.152	.014	-.119	.017	.129	.011	.102
	τ_{00}	**.002**	-.041	.006	-.073	.003	.012	.003	-.053	.003	**.011**	.004	-.056
	τ_{01}	.001	-.036	**.000**	-.008	**.000**	**-.003**	**.000**	-.015	**.000**	-.004	**.000**	-.014
	τ_{11}	**.000**	-.020	.001	-.034	**.000**	**-.005**	.001	-.030	**.000**	-.008	.001	-.033
$\sqrt{0.3}$	β_0	**.003**	.037	.013	-.057	**.003**	**.036**	**.003**	.037	.004	.052	.004	.050
	β_1	.009	-.090	**.006**	**-.049**	**.006**	-.071	**.006**	-.071	.007	-.083	.007	-.082
	σ^2	.008	-.087	.002	.024	**.001**	.027	.017	-.128	**.001**	**-.015**	.008	.090
	τ_{00}	**.002**	-.039	.010	-.095	.003	**-.027**	.007	-.080	.003	-.031	.006	-.077
	τ_{01}	.002	-.040	**.001**	**-.019**	**.001**	**-.019**	**.001**	-.027	**.001**	**-.019**	**.001**	-.021
	τ_{11}	**.000**	-.020	.002	-.044	**.000**	**-.015**	.002	-.042	.001	-.019	.002	-.039
$U(0.55,0.9)$	β_0	.003	.038	.003	-.021	**.002**	**.019**	**.002**	.023	**.002**	.033	**.002**	.030
	β_1	.006	-.075	**.002**	**-.030**	**.002**	-.036	**.002**	-.038	.003	-.045	**.002**	-.044
	σ^2	.008	-.091	**.002**	**-.025**	.033	.179	.013	-.112	.027	.160	.008	.088
	τ_{00}	**.002**	-.043	.004	-.059	.004	**.023**	.003	-.046	.004	**.023**	.003	-.050
	τ_{01}	.001	-.034	**.000**	-.008	.001	.001	**.000**	-.014	.001	**-.000**	**.000**	-.014
	τ_{11}	**.000**	-.020	.001	-.031	**.000**	.007	.001	-.026	**.000**	**.003**	.001	-.028
		N = 2,000											
$\sqrt{0.8}$	β_0	.001	.025	**.000**	**-.016**	.001	.021	.001	.022	.001	.032	.001	.031
	β_1	.004	-.066	**.000**	**-.012**	**.000**	-.020	.001	-.022	.001	-.028	.001	-.026
	σ^2	.011	-.104	**.000**	**.005**	.055	.234	.004	-.064	.049	.222	.004	.061
	τ_{00}	.004	-.062	**.001**	-.031	**.001**	-.028	**.001**	-.024	**.001**	**-.023**	**.001**	-.026
	τ_{01}	.001	-.022	**.000**	**-.001**	**.000**	-.009	**.000**	-.009	**.000**	-.012	**.000**	-.009
	τ_{11}	.001	-.032	**.000**	**-.015**	.001	-.024	**.000**	-.017	.001	-.028	**.000**	-.020
$\sqrt{0.5}$	β_0	**.002**	**.033**	.003	-.038	**.002**	**.033**	**.002**	.035	.003	.046	**.002**	.044
	β_1	.007	-.080	**.001**	**-.028**	.002	-.048	.003	-.049	.003	-.058	.003	-.056
	σ^2	.013	-.114	**.003**	**.050**	.024	.154	.016	-.125	.017	.130	.010	.102
	τ_{00}	.005	-.068	.004	-.061	**.001**	**.004**	.003	-.057	**.001**	.005	.003	-.055
	τ_{01}	.001	-.025	**.000**	-.012	**.000**	**-.001**	**.000**	-.014	**.000**	-.003	**.000**	-.016
	τ_{11}	.001	-.035	.001	-.034	**.000**	**-.004**	.001	-.031	**.000**	-.008	.001	-.032
$\sqrt{0.3}$	β_0	**.001**	**.032**	.011	-.072	.002	.033	.002	.034	.003	.049	.003	.048
	β_1	.009	-.091	**.003**	**-.041**	.005	-.071	.005	-.071	.007	-.083	.007	-.082
	σ^2	.014	-.119	016	.122	.002	.039	.018	-.133	**.000**	**-.005**	.009	.097
	τ_{00}	.005	-.072	.008	-.090	**.002**	**-.038**	.007	-.081	**.002**	-.040	.006	-.077
	τ_{01}	.001	-.029	**.000**	-.017	**.000**	**-.012**	.001	-.024	**.000**	-.013	**.000**	-.020
	τ_{11}	**.001**	-.037	.002	-.044	**.001**	**-.022**	.002	-.042	**.001**	-.024	.002	-.040
$U(0.55,0.9)$	β_0	**.001**	.027	**.001**	**-.023**	**.001**	.029	.002	.032	.002	.043	.002	.040
	β_1	.006	-.073	**.001**	**-.026**	.002	-.043	.002	-.046	.003	-.053	.003	-.050
	σ^2	.013	-.114	**.002**	**.038**	.031	.173	.014	-.116	.024	.150	.009	.095
	τ_{00}	.005	-.067	.004	-.058	.002	**.012**	.003	-.051	**.001**	.015	.003	-.050
	τ_{01}	.001	-.025	**.000**	-.007	**.000**	**.004**	**.000**	-.013	**.000**	**.001**	**.000**	-.014
	τ_{11}	.001	-.034	.001	-.030	**.000**	**.002**	.001	-.027	**.000**	**-.002**	.001	-.029

Note. MCM denotes the moment estimation method. MMLE-U denotes the marginalized MLE when correcting measurement error solely.

Table 5. Mean and covariance matrix of ξ.

	MCMC	EM	MAP
mean	$[1.052 \quad 1.088 \quad 1.156]$	$[1.159 \quad 0.806 \quad 0.904]$	$[1.219 \quad 0.945 \quad 1.064]$
covariance	$\begin{bmatrix} .049 & .058 & .072 \\ .058 & .082 & .106 \\ .072 & .106 & .149 \end{bmatrix}$	$\begin{bmatrix} .030 & .020 & .025 \\ .020 & .035 & .033 \\ .025 & .033 & .062 \end{bmatrix}$	$\begin{bmatrix} .106 & .064 & .072 \\ .064 & .109 & .092 \\ .072 & .092 & .219 \end{bmatrix}$

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS: 88), "Base Year Through Second Follow-up."

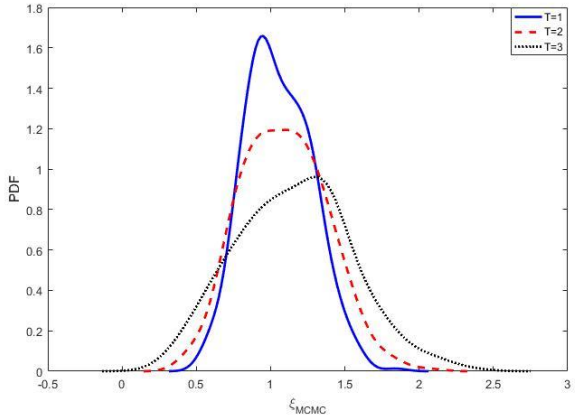

Figure 1. Probability density function of $\hat{\xi}_{MCMC}$ at different time points.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS: 88), "Base Year Through Second Follow-up."

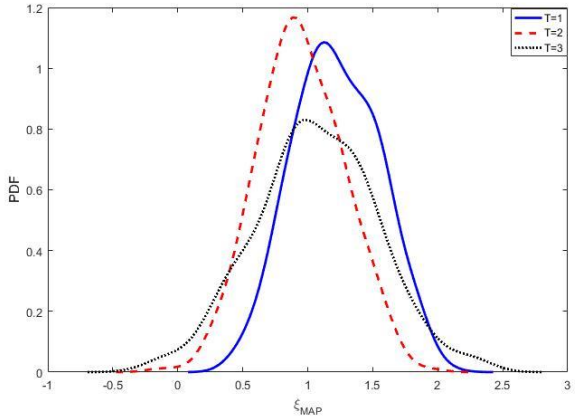

Figure 2. Probability density function of $\hat{\xi}_{MAP}$ at different time points.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS: 88), "Base Year Through Second Follow-up."

Table 6. Parameter estimates and standard errors of NELS 88.

Par.	MCMC	MEM	MMLE	MCEM	MMLE-U	Naïve
β_0	1.046	1.089	1.196	1.135	1.153	1.154
β_1	0.052	-0.127	-0.073	-0.103	-0.077	-0.078
σ^2	0.025	0.047	0.283	0.006	0.270	0.013
τ_{00}	0.066	-0.007	0.175	0.040	0.179	0.036
τ_{01}	0.060	0.022	0.078	0.016	0.073	0.015
τ_{11}	0.026	-0.011	0.036	0.011	0.038	0.008

Note. MEM denotes the moment estimation method. MMLE-U denotes the marginalized MLE when only correcting measurement error.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS: 88), "Base Year Through Second Follow-up."